UK Centre for
Ecology & Hydrology

A Nationally Consistent Approach to Assessing Accumulated Rainfall Rarity

Michael Eastman[1], Simon Parry[1], Cecilia Svensson[1], Jamie Hannaford[1, 2]

[1]UK Centre for Ecology & Hydrology

[2]Irish Climate and Research UnitS (ICARUS), Maynooth University

11/02/2021

Title A Nationally Consistent Approach to Assessing Accumulated Rainfall Rarity

UKCEH contact details National Hydrological Monitoring Program (NHMP)
UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, United Kingdom

t: +44 (0)1491 838800
e: nhmp@ceh.ac.uk

Author Michael Eastman;Simon Parry;Cecilia Svensson;Jamie Hannaford

Approved by Lisa Stewart

Signed EJ Stewart

Date 11/02/2021

Contents

1 Introduction 4
2 Data 6
2.1 Assumptions 7
2.1.1 Autocorrelation 7
2.1.2 Stationarity 11
3 Methods 15
3.1 Distribution Fitting 15
3.1.1 Probability Distributions 15
3.1.2 Parameter estimation 16
3.2 Goodness-of-fit 16
3.2.1 Confidence Intervals 16
3.2.2 Visual goodness-of-fit 17
3.2.3 Goodness-of-fit variation 17
4 Results 18
4.1 Most accepted distributions 18
4.1.1 Overall goodness-of-fit 18
4.1.2 Modified-Anderson-Darling-Lower goodness-of-fit for different accumulation periods, start months and regions 20
4.1.3 12-months or shorter accumulation periods 24
4.1.4 Modified-Anderson-Darling-Lower goodness-of-fit variation – 12-months or shorter accumulation periods 25
4.2 Sensitivity 27
5 Hydrological Summary Case Study 29
6 Summary 36
7 Acknowledgements 38
8 Reference List 39
9 Appendices 41
9.1 Appendix 1 – Autocorrelation plots of Forth accumulated rainfall events beginning in October, varying by accumulation length. The time series consist of overlapping events for durations of 12 months or more. The x-axis shows units of events. 41
9.2 Appendix 2 – Autocorrelation plots of 36-month (overlapping) accumulated rainfall events beginning in October, in each of the 18 regions. The x-axis shows units of events, that is, in this case a 3-year period. 42

1 Introduction

Accumulated rainfall totals are an important variable for a range of hydrological applications, including monitoring and forecasting, and long-term planning.

For more than three decades, the National Hydrological Monitoring Programme (https://nrfa.ceh.ac.uk/nhmp) at UKCEH has produced the monthly Hydrological Summary for the UK. A key component of this publication, a strategically important output with a wide readership, is a table of rainfall accumulations over a range of durations to present, and an assessment of the rarity of wet and dry periods in the context of rainfall data back to 1910. However, there is a need to update the methodology used to derive rainfall return periods, and to assess the rainfall frequencies more accurately, particularly from a low rainfall perspective.

At the same time, for long-term water resources planning there is a requirement to stress test water supply systems against increasingly extreme dry periods. The latest industry trends are focused on 1-in-200-year and 1-in-500-year drought events, a particularly challenging exercise when observational data until recently spanned only around 100 years and, despite recent advances in data rescue and digitisation (Hollis *et al*. 2019), still only cover the last 150 years. Previous approaches to characterising such extreme events have been necessarily pragmatic by their own admission (e.g. the Drought Vulnerability Framework – Counsell *et al.* 2017).

The Drought Libraries developed through the UK Drought & Water Scarcity Programme combined state-of-the-art observational (HadUK, Hollis *et al*. 2019) and simulated (weather@home 2, Guillod *et al.* 2017) datasets into an interactive application to assist water resource managers in evaluating present and future risk from severe droughts. Despite the additional data included, provisional return periods in the Drought Libraries were calculated similarly pragmatically, though recognising the need for further research into more robust methodologies for assessing rarity.

The above applications approach the same question from different perspectives. For the Hydrological Summary, the interest is in calculating the return periods of observed rainfall accumulations. For long-term planning (e.g. Drought Libraries), the focus is on determining the system response that corresponds to certain return periods (e.g. 1-in-200 or 1-in-500 years). Nevertheless, both applications will benefit from a renewed assessment of the frequency estimation methodology. Furthermore, a nationally consistent approach is beneficial for both applications. In the Hydrological Summary, a simple pragmatic approach is sought in which the actual return period value is not stated precisely; instead, an interval is presented to account for uncertainty. For long-term planning, a nationally consistent approach for assessing rainfall rarity will help support the recent shift in focus within the water industry towards regional and national scale planning, through initiatives such as the regional groupings (e.g. Water Resources South East) and the Environment Agency led National Framework. There is a need to adopt consistent approaches in order to tackle the important questions of the future, such as inter-regional water transfers and enhanced resilience. While the latest round of guidance specifies resilience to extreme droughts (e.g. 1-in-500 years) should be assessed in terms of water supply system metrics rather than hydrometeorological inputs, there remains a fundamental need for consistent approaches to indexing extreme rainfall deficits, to understand the propagation of drought severity from rainfall inputs through runoff/recharge to system response (Environment Agency, 2020).

The aim of this report is to identify the most appropriate approach to quantifying return periods of long duration rainfall. The Data and Methods sections describe the data used to test the range of distributions and goodness-of-fit measures. Results are presented which identify the most appropriate distribution for nationally consistent applications, before the key findings are summarised.

2 Data

This study uses the HadUK gridded rainfall dataset (Hollis *et al*. 2019), which represents the state of the art in observed, nationally consistent rainfall data. Time series of rainfall from 1862 to 2017 for 18 regions (Figure ***1***) were extracted by averaging gridded rainfall over the spatial domains. Rainfall data were accumulated over varying periods, producing a range of time series. The final set of accumulated rainfall time series encompassed one series for each combination of the 18 regions, 13 accumulation periods (1, 2, 3, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60 months), and 12 start months (Jan-Dec).

Figure 1 *Regions over which HadUK gridded rainfall was averaged.*

2.1 Assumptions

The parameter estimation of theoretical distributions depends on a number of assumptions, and therefore the power of hypothesis tests to evaluate the goodness-of-fit of a distribution, and the estimated return periods themselves, also depend on these assumptions. The assumptions that events are independent and identically distributed, i.e. that data are independent from each other and of time, are explored further.

2.1.1 Autocorrelation

Autocorrelation, also referred to as serial dependence, is the correlation of events with prior events, suggesting the events do not occur randomly. Figure ***2*** shows some examples of how autocorrelation varies within the HadUK accumulated rainfall time series (Venables and Ripley, 2002; R Core Team, 2019). The degree of autocorrelation does not appear to be dependent on the start month. However, autocorrelation does vary regionally and by accumulation period length, as discussed further below.

Figure 2 demonstrates the increasing presence of autocorrelation within the rainfall series as the accumulation period increases, with a more elaborate example given in Appendix 1. This serial dependence is to be expected, as accumulated periods of more than 12 months overlap, and thus are not independent. However, the presence of autocorrelation also in accumulation durations of 12 months or shorter shows that overlapping is not the only source of autocorrelation.

Whilst evidence of autocorrelation exists in all regions, the extent of autocorrelation is higher in Forth and Welsh than in the other regions presented in Figure ***2***. Appendix 2 shows the autocorrelation across all the analysed regions for 36-month accumulated rainfall beginning in October. Autocorrelation appears to be more common in the north and west of the UK. This is possibly due to the relationship between long-period, large-scale atmospheric circulation, such as the North Atlantic Oscillation (NAO), and rainfall in the North West (e.g. Svensson *et al.* 2015).

To minimise autocorrelation in the time series, distributions were fitted to non-overlapping data. This was achieved by extracting multiple time series from accumulation periods exceeding 12 months. The start month was lagged by a year iteratively until the start month matched the start of the second accumulation period in the original time series (Figure ***3***). This produced a number of non-overlapping datasets, the number and size of which are demonstrated in Table 1. Estimates derived from the frequency curve, such as return period and return level, were subsequently averaged over the multiple non-overlapping time series to produce single estimates.

Comparisons of Figure 2 and Figure 4 demonstrate the considerable reduction in autocorrelation caused by removing overlapping data from the time series. Whilst detectable, the extent of autocorrelation is minimal across different lengths accumulation period, start months and regions.

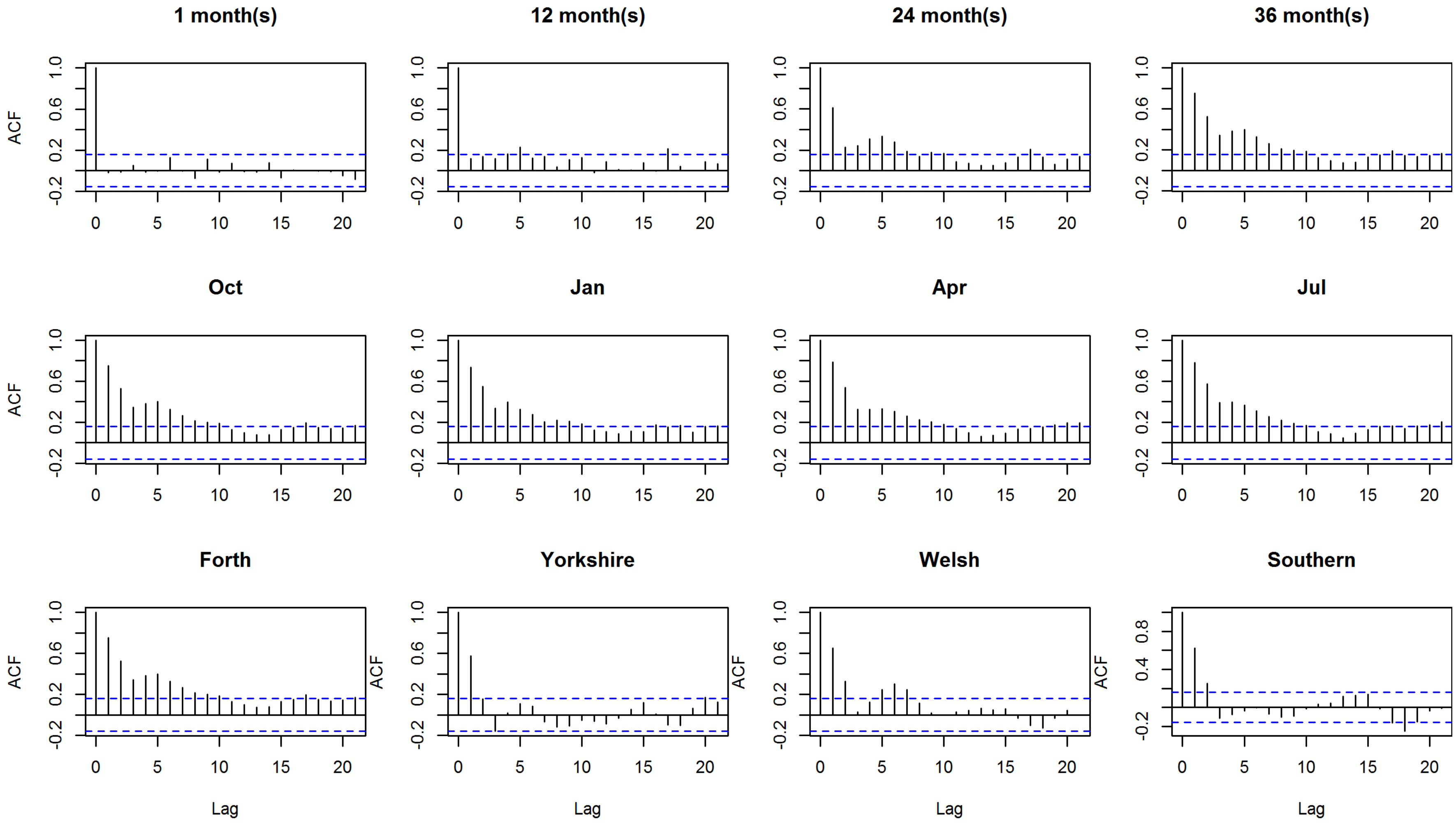


Figure 2 *Autocorrelation plots of overlapping accumulated rainfall time series demonstrating how autocorrelation varies by; accumulation period in Forth, beginning in October (1st row), start month in 36-month accumulations in Forth (2nd row), and region in 36-month accumulations beginning in October (3rd row).*

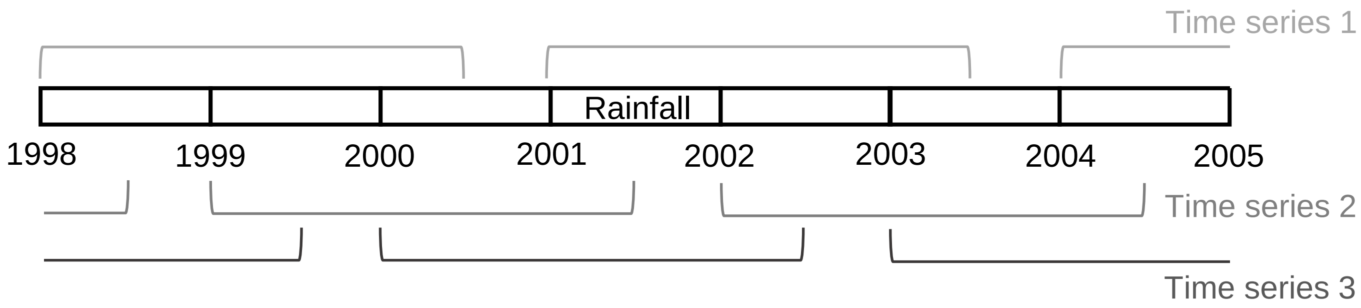


Figure 3 *Time series sampling method. This example shows the data included in each of the three non-overlapping time series that can be extracted for the 30-month duration.*

Table 1 *The number of non-overlapping monthly-averaged rainfall data included in each extracted time series.*

Accumulation Period (months)	**Number of time series**	**Number of monthly averages in each time series (TS)**				
		TS 1	TS 2	TS 3	TS 4	TS 5
1	1	156	-	-	-	-
2	1	156	-	-	-	-
3	1	156	-	-	-	-
6	1	156	-	-	-	-
12	1	156	-	-	-	-
18	2	78	77	-	-	-
24	2	78	77	-	-	-
30	3	52	51	51	-	-
36	3	52	51	51	-	-
42	4	39	38	38	38	-
48	4	39	38	38	38	-
54	5	31	31	30	30	30
60	5	31	31	30	30	30

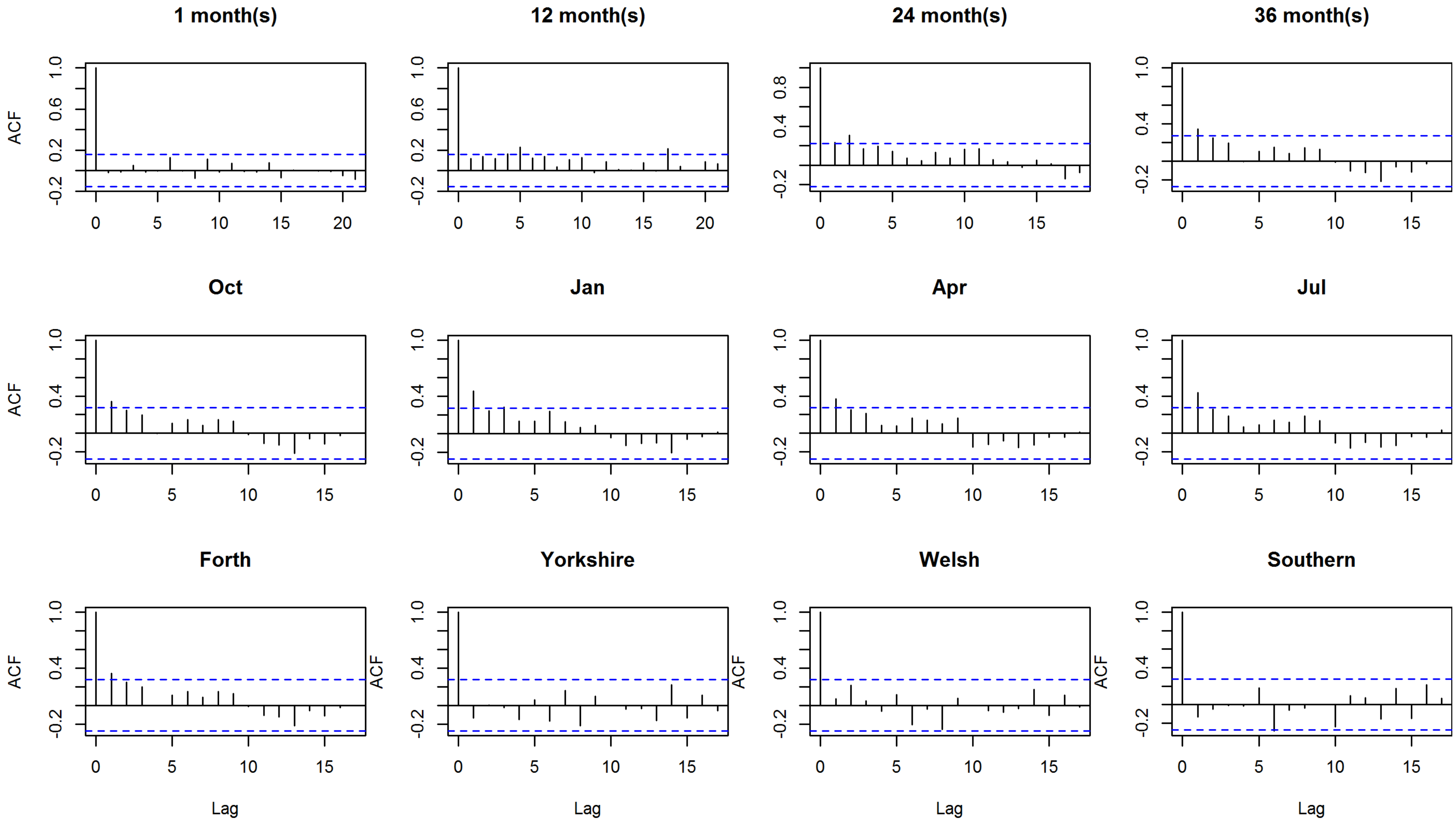


Figure 4 *Autocorrelation plots of non-overlapping accumulated rainfall time series demonstrating how autocorrelation varies by; accumulation period in Forth, beginning in October (1st row), start month in 36-month accumulations in Forth (2nd row), and region in 36-month accumulations beginning in October (3rd row)*

2.1.2 Stationarity

The second assumption investigated is whether data are identically distributed. In hydrological data this assumption is typically breached by data showing seasonality, or by there being a temporal trend in the data. Here, we extract data series of fixed durations, starting at the same time of year for all the values in the same series, thus ensuring that all the data in a particular series have the same seasonal characteristics and thus are identically distributed with regards to season. If instead, say, 1-month duration rainfalls had been extracted regardless of season, then the data for months in the rainy season might tend to be larger than the data for months in the dry season, making the data series non-identically distributed over time.

Another aspect of non-stationarity is when data increase or decrease over longer time periods, say, because of climate change or because of long-period oscillations such as the North Atlantic Oscillation (NAO). The dependence on time can also show up as serial dependence in data if trends are reasonably pronounced. In this section we explore if there are any temporal trends in the extracted non-overlapping data series.

Figure ***5*** shows examples of how accumulated rainfall changes with time. In general, these examples demonstrate little visual evidence of trends in the rainfall data.

To further explore possible trends in the data, Mann-Kendall trend statistics and their significance were estimated for time series beginning in October and in March using the Kendall v2.2 R package (Mann 1945; McLeod 2011). The first non-overlapping time series was extracted from accumulation periods exceeding 12 months according to the process outlined above. The full time series was used for time series with up to 12-month accumulation periods, as they do not overlap and were therefore the extraction of multiple non-overlapping time series was not necessary. This enabled the general presence/absence of trends to be assessed, as well as variation with accumulation period length and region. The results are shown in Figure 6 and Figure 7. Statistically significant trends (indicated by an asterisk) were identified across all accumulation periods. However, trends are more prominent in rainfall time series with longer accumulation periods. The presence of trends also appears to vary spatially, with statistically significant trends being more common in the North and West of the United Kingdom. This agrees with the autocorrelation analysis, which was concentrated in the North West, and again suggests a link with the NAO which mainly affects rainfall in this part of the country (e.g. Svensson *et al.* 2015).

Trends for accumulation periods beginning in March and October demonstrate similar variation in significance across accumulation periods and regions. Both exhibit higher frequency of significant trends in the North and West, and for longer accumulation periods. Significant trends were more common in rainfall accumulation periods beginning in October (87/234) than March (68/234).

Despite the presence of trends in the data, and the potential for this to influence the fitted statistical distributions, developing a non-stationary frequency model was beyond the scope of this study. Therefore, the temporal trends in the North and West in winter are ignored, with the caveat that frequency estimates for this part of the country may not be representative of a future period. The development of non-stationary drought frequency models remains an important topic for further research.

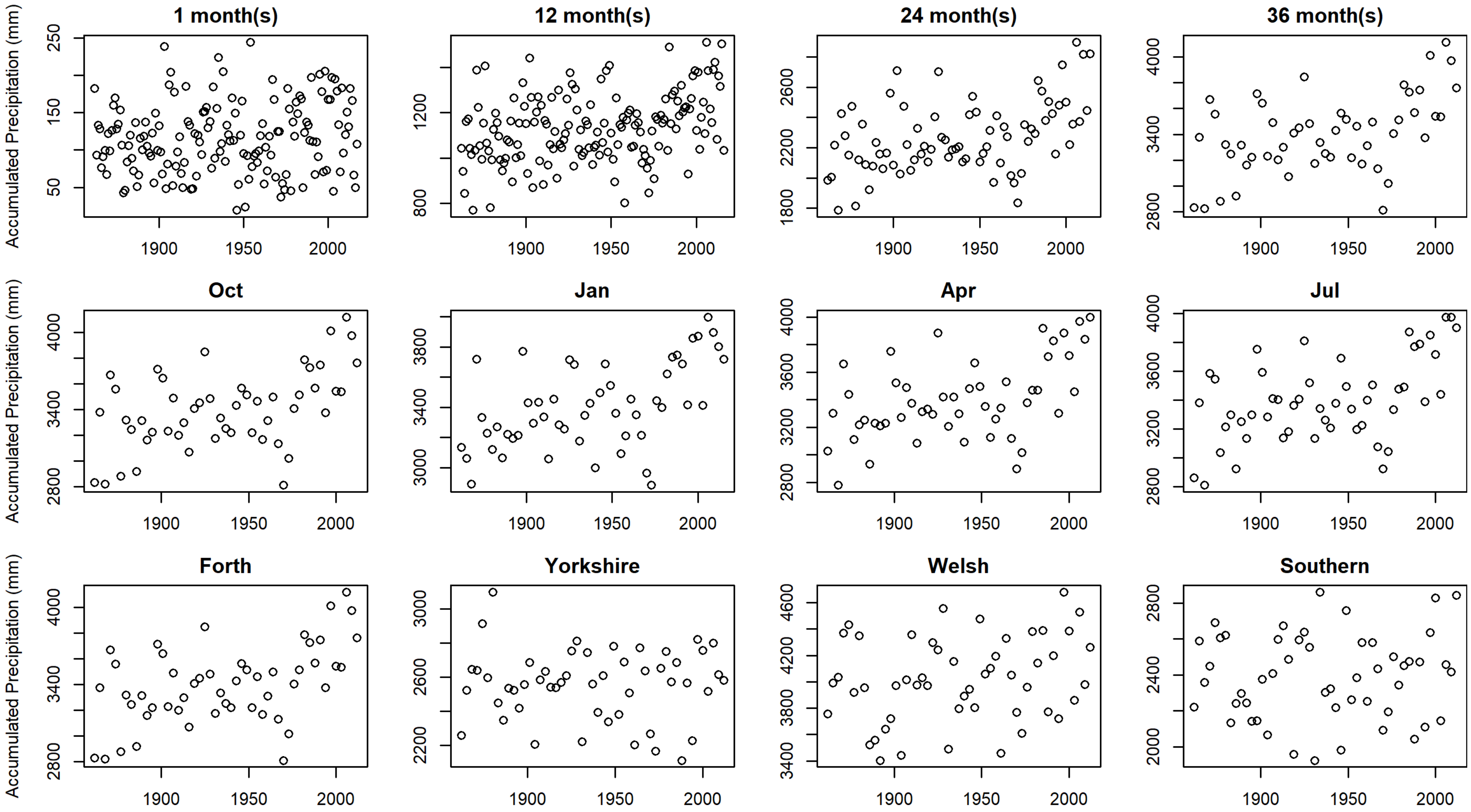


Figure 5 *Non-overlapping accumulated rainfall time series plotted against time demonstrating variation by; accumulation period in Forth rainfall beginning in October (1st row), start month in 36-month accumulated rainfall in Forth (2nd row), and region in 36-month rainfall beginning in October (3rd row).*

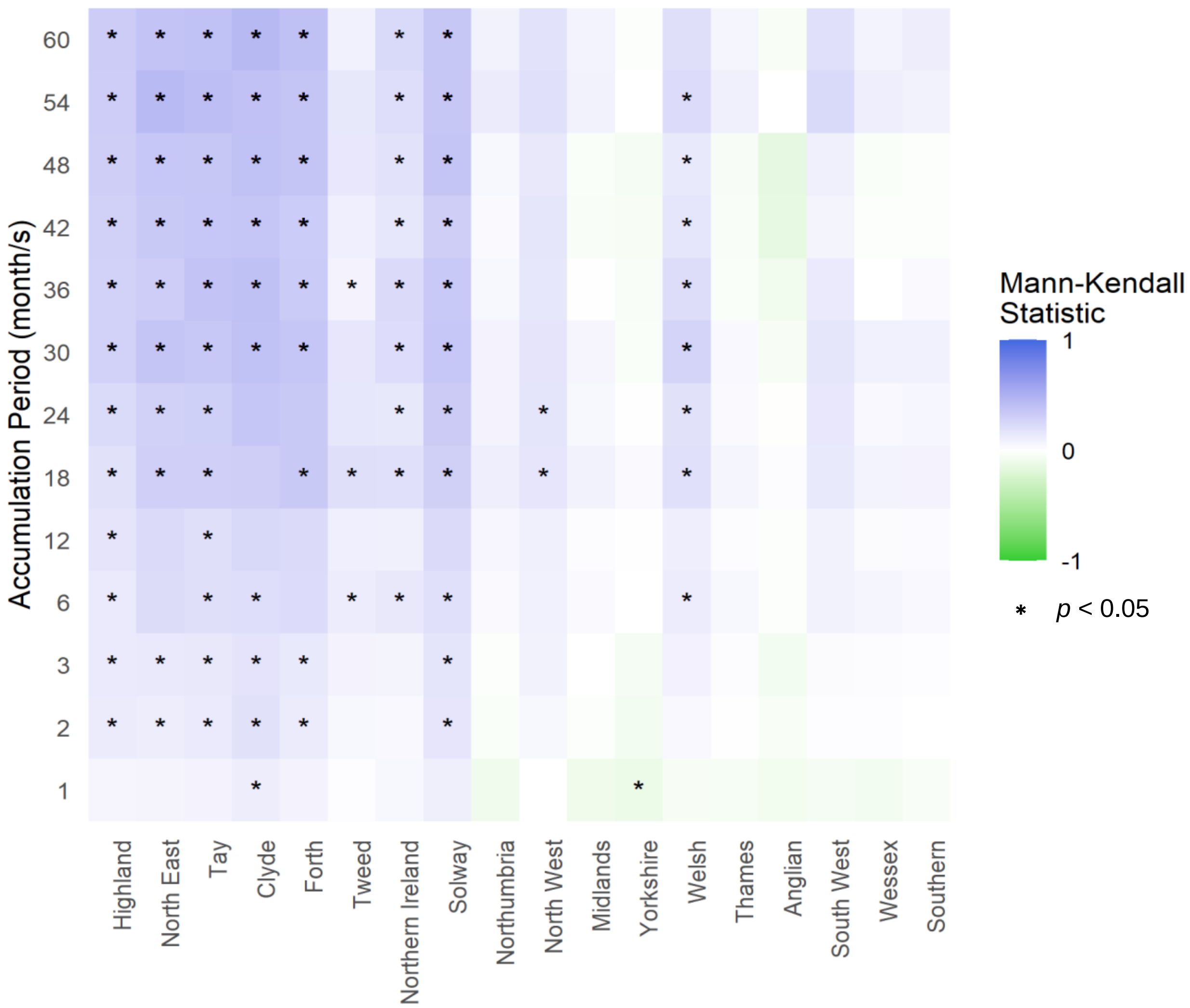


Figure 6 *Trends in accumulated rainfall beginning in October across accumulation periods and regions.*

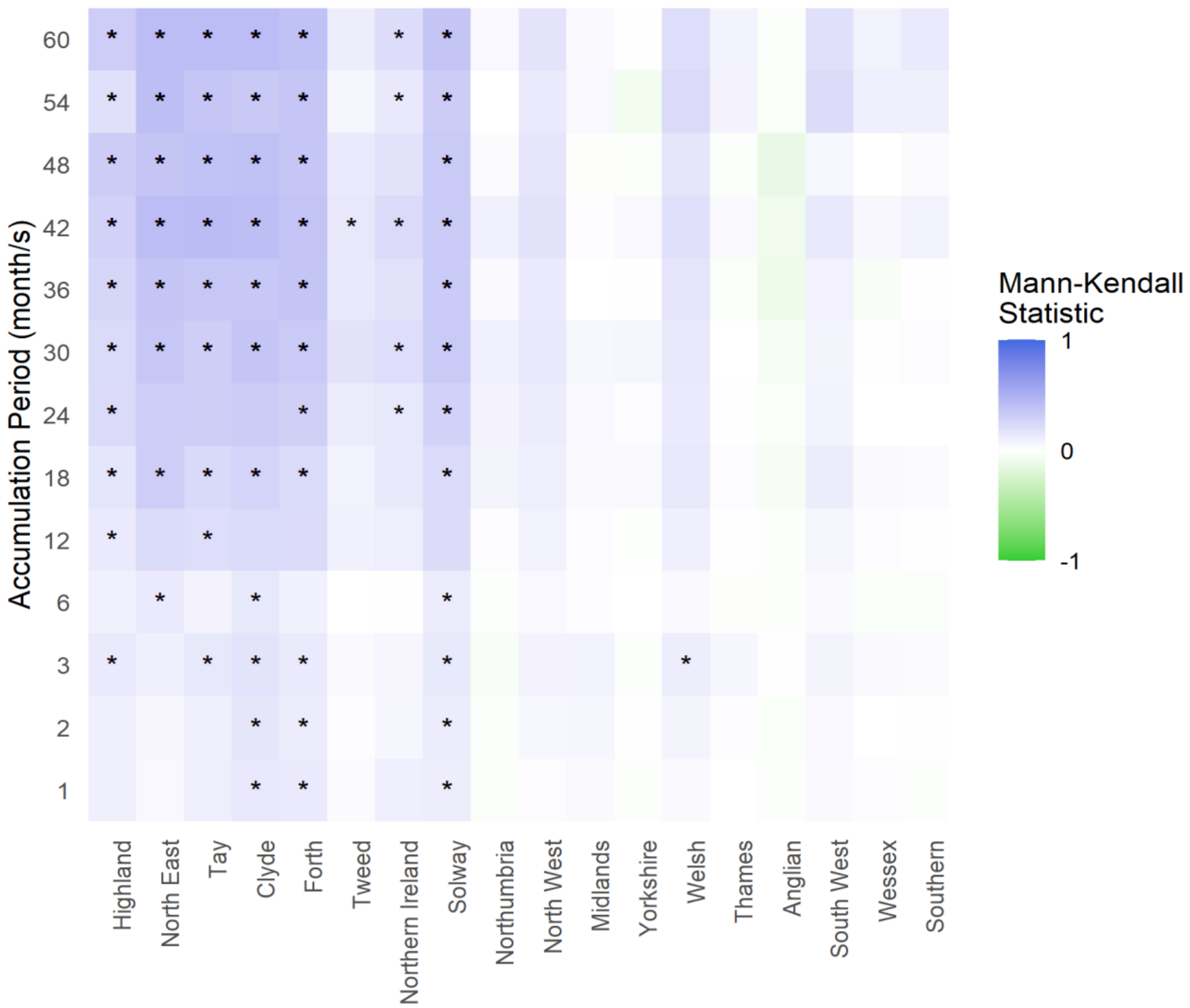


Figure 7 *Trends in accumulated rainfall beginning in March across accumulation periods and regions.*

3 Methods

3.1 Distribution Fitting

3.1.1 Probability Distributions

Extreme value analysis enables the estimation of rare events using extreme value distributions. Annual Maximum/minimum and peak-over-threshold (POT) approaches are common approaches to rare event probability estimation. However, both methods are sensitive to sample size, and in cases where data are limiting, other approaches may be deemed a more robust. In the present study, the HadUK gridded rainfall dataset contains 156 years of data. However, with the subsetting of the data in order to avoid serial dependence, the sample sizes can become as low as 31. Therefore, other distributions are also investigated here, and compared with the distributions in the extreme value distribution families.

The suitability of nine probability distributions was evaluated; Gaussian, Generalised Gaussian, Gamma, Generalised Extreme Value (GEV), Generalised Logistic, Gumbel, Log-Gaussian, Pearson Type III and Weibull. Of these, the GEV, Gumbel, and Weibull are extreme value distributions (Table 2).

The Gaussian, Generalised Gaussian, Pearson Type III, Generalised Logistic, Generalised Extreme Value and Gumbel distributions do not have a lower bound at zero, unlike observed precipitation. However, because a single month's accumulation of rainfall in the UK rarely approaches zero, these probability distributions were included in the analysis despite their lack of a lower bound at zero.

Table 2 *Statistical distributions investigated.*

Distribution	**Abbreviation**	**Extreme value distribution**	**R package**	**Reference**
Gaussian	gau		stats	R Core Team, 2019
Generalised Gaussian	gno		gnorm	Griffin, 2018
Gamma	gam		stats	R Core Team, 2019
Generalised Extreme Value	gev	✓	evd	Stephenson, 2002
Generalised Logistic	glog		SCI	Gudmundsson & Stagge, 2016
Gumbel	gum	✓	evd	Stephenson, 2002
Log-Gaussian	lno		stats	R Core Team, 2019
Pearson Type III	pe3		SCI	Gudmundsson & Stagge, 2016
Weibull	wei	✓	stats	R Core Team, 2019

3.1.2 Parameter estimation

The maximum likelihood estimation (MLE) method was used to estimate the probability distribution parameters. MLE was bounded where necessary. Starting values were generated using L-moments for all distributions using the 'lmom' R package. L-moment estimations of zero-bounded parameters that were below zero were set to 0.1.

3.2 Goodness-of-fit

Goodness-of-fit tests are often used to identify the most appropriate statistical probability distribution for a given empirical frequency distribution. These tests enable the agreement between the observed and assumed distributions to be assessed. Although a number of tests exist, the Anderson-Darling test is particularly appropriate since it utilises a weight function that emphasises agreement between the observed and assumed probability distribution in the tails (Anderson and Darling 1952).

The Anderson-Darling test assigns equal weight to both tails. However, since the focus of this report is on dry weather events, a modification of the Anderson-Darling test was also used to determine goodness-of-fit (Sinclair *et al.* 1987; Sinclair *et al.* 1990). This modification gives higher weighting to the lower tail.

3.2.1 Confidence Intervals

Estimation of uncertainty in the parameter estimates, and therefore (original and in the Modified) Anderson-Darling test statistics is required in order to determine whether the theoretical distribution fits the empirical distribution. Parametric bootstrapping is a regularly implemented method in which the uncertainty of distribution parameters can be estimated (Kharin and Zwiers 2000; Paeth and Hense 2005; Kyselý 2007).

Parametric bootstrapping was applied here by extracting *n* random values 10,000 times from the probability distributions, where *n* refers to the sample size of the empirical distribution. Theoretical distribution parameters were estimated according to the method outlined in Section 3.1.2, and a goodness-of-fit score (the Anderson-Darling test statistic) was calculated for each of the 10,000 bootstrapped samples. The 5^{th} and 95^{th} quantiles of these goodness-of-fit scores were then extracted, and this 90% confidence interval was used to determine whether or not the theoretical distribution fitted the empirical distribution. Hence, the null hypothesis that the observed empirical distribution comes from the theoretical distribution can/cannot be rejected at the 10% significance level. The null hypothesis was accepted in cases where the goodness-of-fit score derived from the observed, empirical distribution falls within the 90% confidence interval.

Accumulation periods exceeding 12 months

For 1 to 12 month accumulation periods the parametric bootstrapping method outlined in Section 3.2.1 was performed on a single time series, extracted according to the method outlined in Section 2.1.1. However, as multiple time series were extracted from accumulation periods exceeding 12 months to prevent overlapping, each time series was bootstrapped separately. This produced multiple 90% confidence interval bands and thus goodness-of-fit of each extracted, non-overlapping time series was evaluated separately.

Using multiple non-overlapping series means that the number of goodness-of-fit tests carried out will be greater than one for accumulation periods exceeding 12 months. Therefore the proportion, rather than the total number, of null hypotheses accepted was used. The sum of proportions of null hypotheses accepted, and how they vary by accumulation period, start month, and region are explored throughout the remainder of this report to give an indication of the suitability of each distribution for describing the behaviour of accumulated rainfall in the UK. This will be referred to as the null acceptance score for the rest of this report.

3.2.2 Visual goodness-of-fit

Goodness-of-fit was also assessed visually to provide insights into the limitations of certain distributions, as well as to identify systematic differences in goodness-of-fit. This included plotting empirical versus theoretical quantiles, probability density functions superposed on histograms, and return levels versus return periods.

3.2.3 Goodness-of-fit variation

Sources of variation in goodness-of-fit was investigated for all distributions to enable the investigation into how well different distributions fit different types of data. The effect of accumulation period, region, and starting month was determined by comparing the number of null hypotheses accepted for each.

4 Results

4.1 Most accepted distributions

4.1.1 Overall goodness-of-fit

Anderson-Darling

The Anderson-Darling goodness-of-fit test puts equal weight on both the upper and lower tails when evaluating how well a theoretical distribution fits the data sample. The Pearson Type III distribution had an Anderson-Darling null acceptance score of 2,685 out of a possible 2,808, when considering all accumulation periods, start months, and regions. This score was higher than for any other distribution. However, Figure ***8*** demonstrates the relatively high Anderson-Darling null acceptance scores for several of the distributions. The Gamma distribution had the second highest score, 2,556.15. In contrast, the Generalised Gaussian, Weibull and Gumbel distributions had very low null acceptance scores of 1,827.62, 1,631.98, and 1,791.85, respectively.

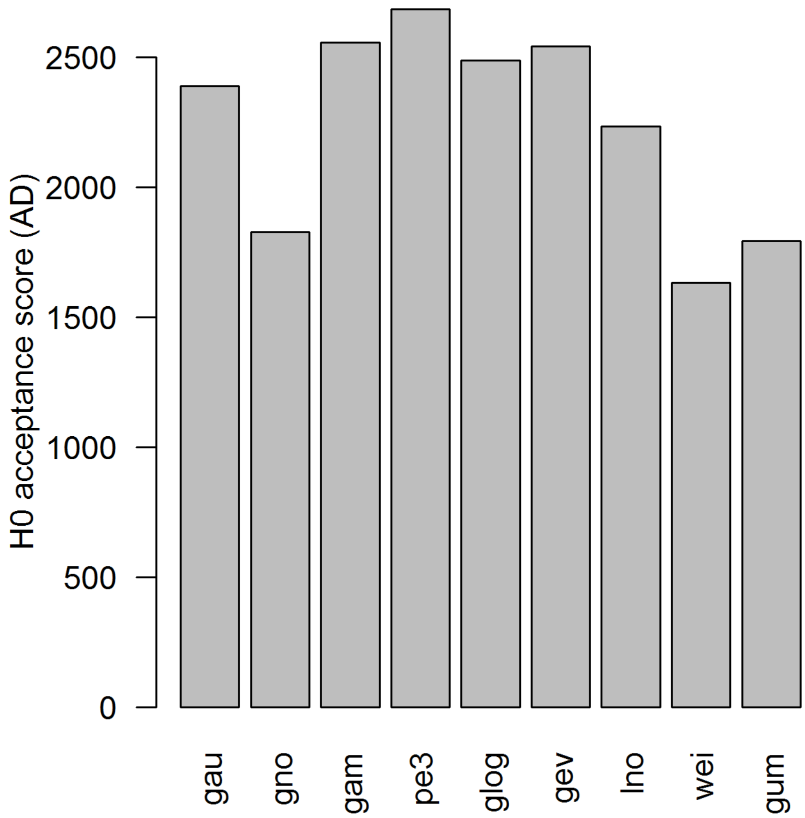


Figure 8 *The sum of proportions of accepted distributions across all combinations of accumulation period, start month, and region, based on the Anderson-Darling goodness-of-fit test.*

Modified-Anderson-Darling-Lower

The focus of this report is on low rainfall events. Therefore, the Modified-Anderson-Darling-Lower test is the primary focus of the results section, as this test puts more weight on the lower than on the upper tail of the distribution. Anderson-Darling results are not presented in the remainder of this report due to the considerable agreement between Anderson-Darling and Modified-Anderson-Darling-Lower results (Figure ***8***; Figure ***9***).

Instead, the remainder of this report documents the comparison of each distribution's Modified-Anderson-Darling-Lower null acceptance scores, the sum of proportions of null hypotheses accepted for each combination of accumulation period, start month, and region, to determine the relative suitability of each distribution for estimating the frequency of occurrence of UK accumulated rainfall events.

The Modified-Anderson-Darling-Lower null acceptance score was highest for the Pearson Type III distribution (Figure ***9***). The Pearson Type III had a null acceptance score of 2,694, out of a possible 2,808. The Gamma distribution had the second highest score, 2,572.

Figure ***9*** demonstrates the relatively poor overall suitability of the Generalised Gaussian and Weibull distributions for estimating return periods of accumulated rainfall events in the UK. The Generalised Gaussian and Weibull distributions had associated scores of 1,822 and 1,611, respectively.

The results presented in Figure ***9*** suggest that a number of distributions may be suitable for low rainfall event return period estimation in the UK. The Modified-Anderson-Darling-Lower null acceptance scores for Gamma, Pearson Type III, Generalised Logistic, and Generalised Extreme Value distributions varied from 2,508-2,694 out of 2.808.

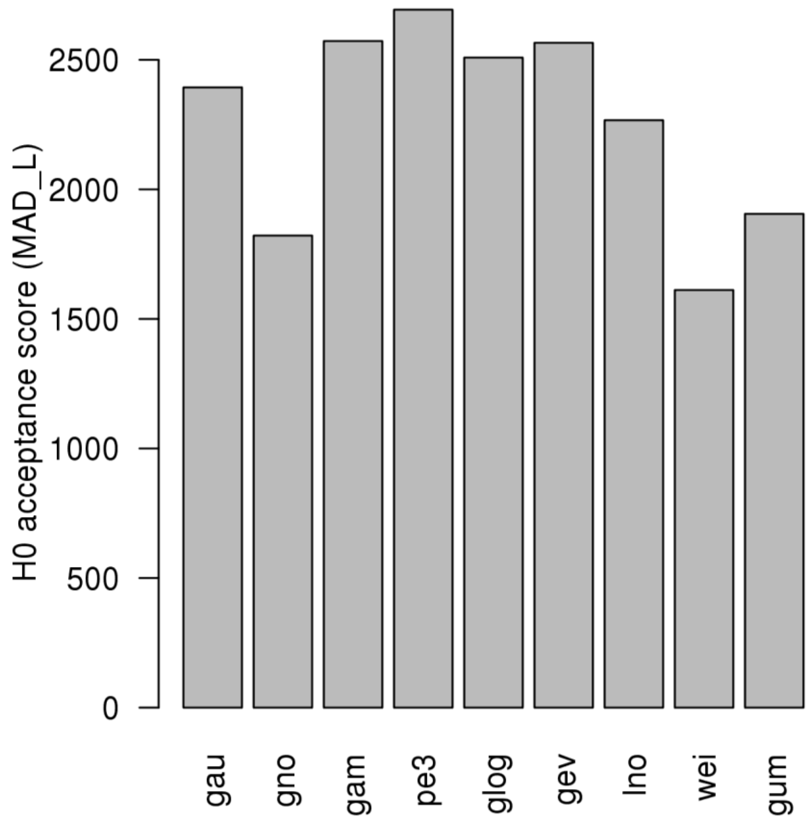


Figure 9 *The sum of proportions of accepted distributions across all combinations of accumulation period, start month, and region, based on the Modified-Anderson-Darling-Lower goodness-of-fit test.*

4.1.2 Modified-Anderson-Darling-Lower goodness-of-fit for different accumulation periods, start months and regions

The observed (empirical) rainfall distributions vary according to the accumulation period, start month, and region they come from (Figure ***10***). The degree of skewness varies between time series, such as 1-month accumulated October rainfall from the Southern region, which was more skewed towards the left than 12-month October-September rainfall from Southern. There is also variation in kurtosis, with 12-month October-September rainfall from Yorkshire exhibiting lighter tails than 12-month October-September rainfall from the Welsh region.

The variation in observed distributions results in varying degrees of null acceptance scores of the fitted theoretical distributions according to accumulation period, region, and start month (Figure ***11***).

Accumulation Period

The Pearson Type III distribution had the highest, or equal highest, null acceptance scores for 6 of the 13 accumulation periods. The GEV had the second highest score across accumulation periods, having the highest, or equal highest, score for 3 accumulation periods.

Pearson Type III was the most appropriate distribution for 1, 2, 3, 6, 12, and 24 month accumulation periods. Although Pearson Type III did not have the highest null acceptance score for other accumulation periods, the score difference between Pearson Type III and the highest scoring distribution was 5 or less for all accumulation periods excluding 54- and 60-months. The Generalised logistic distribution had the highest null acceptance score for the 54- and 60-month accumulation periods, 214.8 and 215 out of a possible 216. Pearson Type III had a null acceptance score of 203.8 and 209.2 for 54- and 60-month accumulation periods, respectively.

Start month

Pearson Type III had the highest, or equal highest, null acceptance score for all start months, excluding June. Gamma had a null acceptance score of 225.45 for time series beginning in June. Whilst this was higher than other distributions, the difference between Gamma and Pearson Type III scores was less than 0.5, as Pearson Type III had a score of 225.0. No other distribution had the highest, or equal highest, null acceptance score for any start month.

Region

Regionally, the results also suggest Pearson Type III is the most appropriate distribution for accumulated rainfall approximation in the UK. Pearson Type III had the highest, or equal highest, null acceptance score for 15 of 18 regions. Gamma had the highest score in Yorkshire, Anglian, and Forth. The Gamma scores for these regions were 147.3, 149.8 and 147.3, respectively. Pearson Type III had associated score of 146.8, 148.3 and 146.1. Therefore, whilst Pearson Type III did not have the highest score for these regions, it was 0.34%, 1.05% and 0.78% lower than the respective Gamma scores.

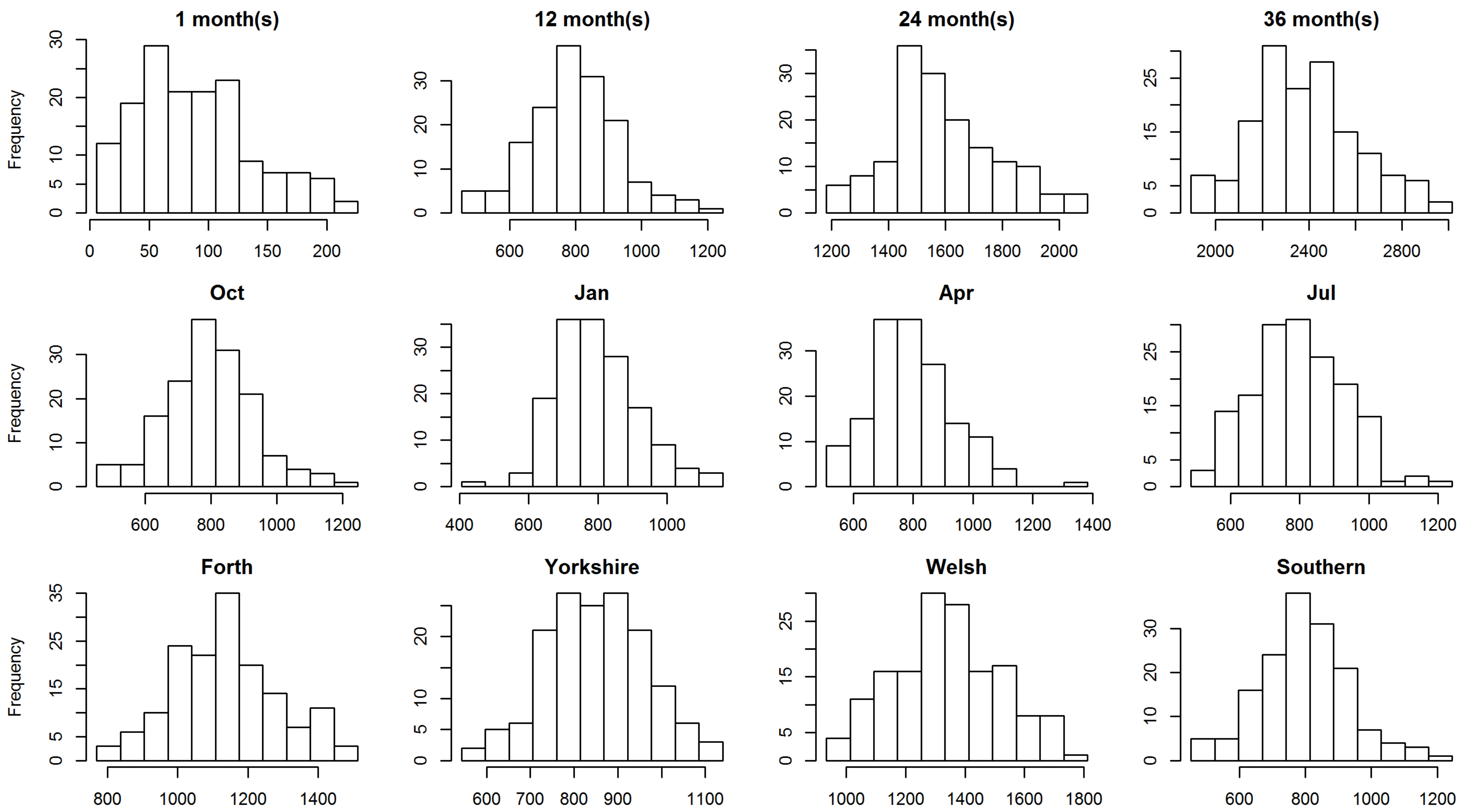


Figure 10 *Examples of observed distributions of rainfall for different accumulation periods in Southern Region rainfall beginning in October (1st row), start months in 12-month accumulated rainfall in Southern (2nd row), and region in 12-month rainfall beginning in October (3rd row).*

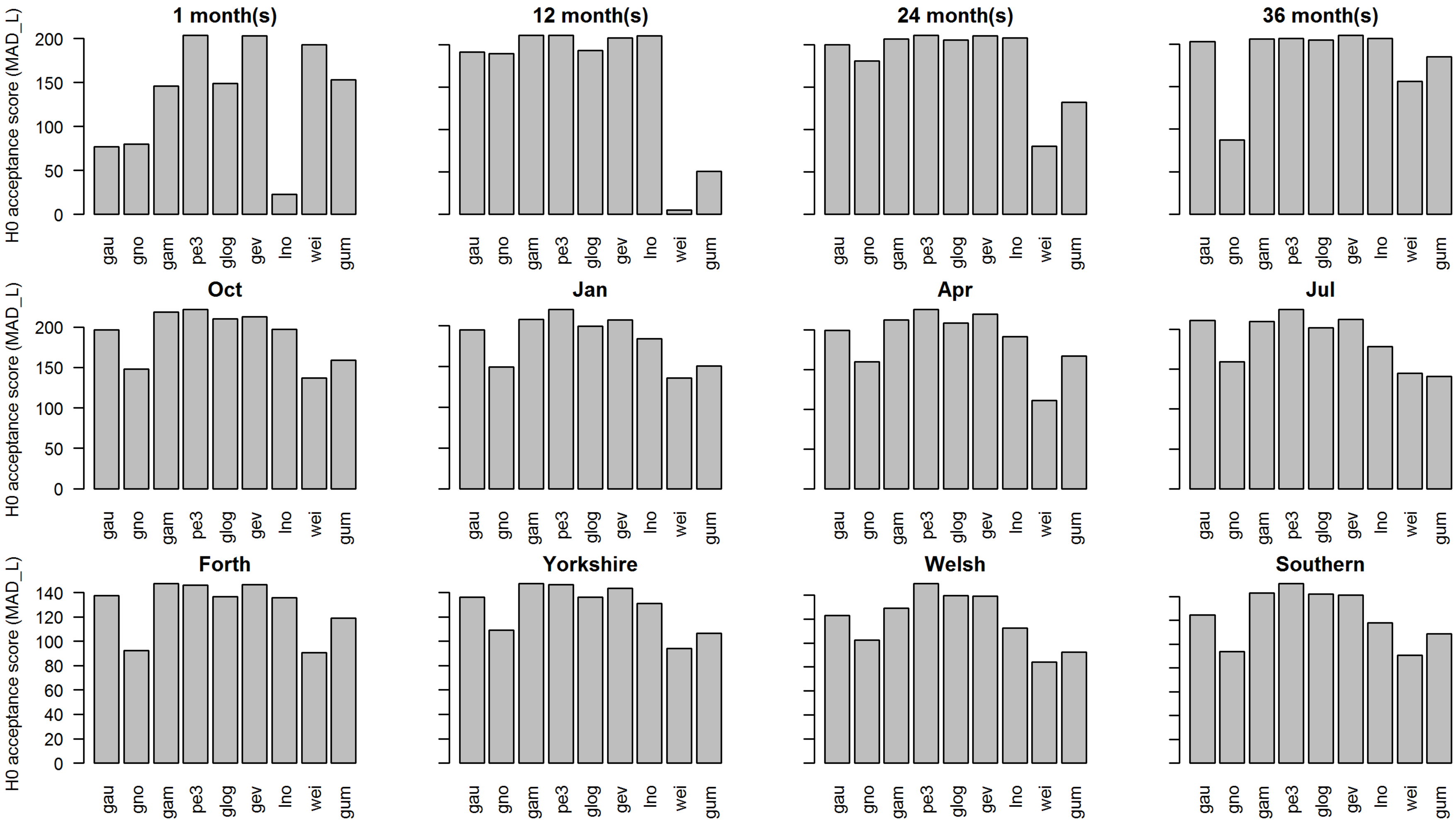


Figure 11 *Null acceptance score for each distribution for different accumulation periods, start months, and regions.*

4.1.3 12-months or shorter accumulation periods

Section 4.1 demonstrates the suitability of the Pearson Type III distribution for estimating return periods and levels when considering all the time series together, as well as across accumulation periods, start months and regions (Figure 8; Figure 9; Figure 11). However, shorter accumulation periods typically exhibit higher skewness than longer durations due to the lower variation exhibited in longer accumulation periods. Therefore, a closer look was taken at the performance of distributions at accumulation periods of 12-months or less. This included 1-, 2-, 3-, 6-, and 12-month accumulation periods.

Generally, Pearson Type III appears to be the most suitable distribution for accumulation periods of 12 months or shorter. According to the Modified-Anderson-Darling (lower) test, 96.5% (1042/1080) of observed time series are acceptably described by a Pearson Type III distribution (Figure ***12***). The GEV also performed well in terms of the Modified-Anderson-Darling-Lower test, with 95.2% of null hypotheses accepted.

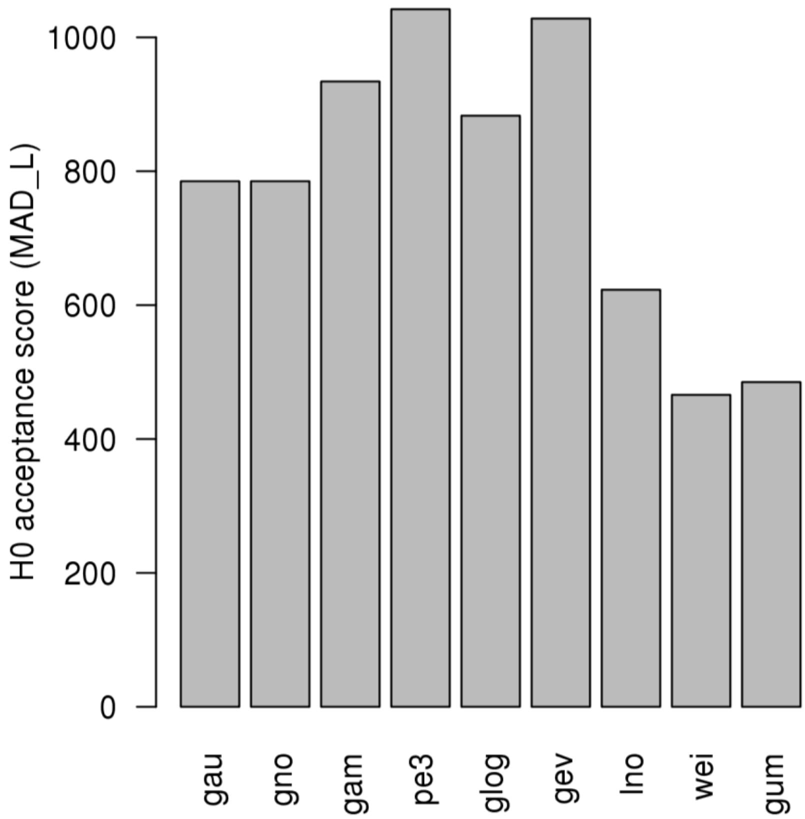


Figure 12 *Total numbers of distribution fits to rainfall accumulation (1, 2, 3, 6, and 12 months) time series that were not rejected by the Modified-Anderson-Darling-Lower test.*

4.1.4 Modified-Anderson-Darling-Lower goodness-of-fit variation – 12-months or shorter accumulation periods

As discussed above, and presented in Figure ***10***, the observed rainfall distributions vary according to the accumulation period, start month and region. This variation drives differences in the ability of distributions to fit the observed distribution, according to accumulation period, start month and region.

Accumulation Period

Pearson Type III was the most accepted distribution for all accumulation periods. GEV null hypotheses were accepted less than the Pearson Type III and Gamma for all accumulation periods. Gamma, however, was accepted for 211 of 216 12-month accumulated rainfall time series, equal to the number of Pearson Type III distributions accepted.

Start month

The Pearson Type III distribution was the most accepted, or equally accepted, in all start months except April and June. The GEV null hypotheses were accepted for 89 of the 90 time series beginning in April and June, making it the most accepted distribution for these two months. The Pearson Type III null hypotheses were accepted in 88 time series for both April and June.

Region

The null acceptance score was higher for the Pearson Type III than any for any other distribution in 13 of the 18 regions. The GEV was the most accepted distribution in Yorkshire, South West, North East, Solway, and Northern Ireland. However, the number of accepted GEV null hypotheses only exceeded that of the Pearson Type III distribution with 1 in Yorkshire, South West, North East and Solway. The difference in Northern Ireland was 2.

Modified-Anderson-Darling-Lower goodness-of-fit variation – 1-3-month accumulation periods

As previously mentioned, this analysis includes a number of unbounded distributions. Neither the Pearson Type III nor the GEV distribution have a lower bound at zero. Whilst this is not an issue for long accumulation periods where rainfall totals are never even close to zero in the UK, fitted theoretical distributions for short accumulation periods exhibiting left skewness or leptokurtosis may produce negative return level estimates. Therefore, unbounded distributions were deemed inappropriate for 1-3-month accumulation periods, and not assessed.

Parametric bootstrapping was performed to determine if unbounded distributions were appropriate for estimating 1-3-month accumulated rainfall return periods. The 25th percentile (the value which is exceeded in ~75% of cases) of the 500-year return level for every combination of start month and region was analysed for 2- and 3-month accumulation periods. The 25th percentile was below zero in 9 of 216 2-month accumulation periods. However, only one 3-month accumulation period time series produced a 500-year return level interquartile range that crossed 0. The 500-year return period for 3-month accumulated rainfall beginning in March, in Southern had a 25th percentile of -0.01mm. Therefore, the above analysis was re-run for the Gamma, Log-Gaussian and Weibull distributions for 1-3-month accumulation periods.

The Gamma distribution was accepted by the Modified-Anderson-Darling-Lower test in 520 of 648 (80.2%) cases of 1-3-month accumulation period time series. This is more than for the Weibull (419) and Log Gaussian (239) distributions.

Whilst the Pearson Type III distribution exhibits a higher Modified-Anderson-Darling-Lower null acceptance score than Gamma, it was deemed inappropriate due to its unbound nature, and therefore not reported on further.

The support for the Gamma distribution is consistent across the regions and accumulation periods. The Gamma distribution was the most accepted distribution for 2 of the 3 accumulation periods, 10 of the 12 start months, and 16 of the 18 regions. The Weibull distribution null hypotheses were accepted more than the Gamma distribution null hypotheses for the 1-month accumulation period, accumulation periods beginning in July and August, and accumulation period time series in Thames, and Wessex.

4.2 Sensitivity

The null hypothesis that the observed distribution comes from a Pearson Type III distribution was accepted for 12-month accumulated rainfall in the Northumbrian region, starting in October. However, the GEV null hypothesis was rejected, suggesting the rainfall time series does not come from a GEV distribution. Figure ***13*** demonstrates the fits of both distributions. The sensitivity of return period estimates to the choice of distribution is evidenced in Figure ***14***.

According to Pearson Type III, a 12-month rainfall event in the Northumbrian region, starting in October, with a spatial average rainfall of 48.09mm is a 1 in 250 year event. However, an average of 48.09mm is associated with a 1 in 559.5 year event according to the GEV distribution. Average rainfall of 45.83mm and 44.56mm is associated with 500 and 750 year return periods, respectively, according to Pearson Type III. However, according to GEV, these events would be 1 in 1568 and 2953 year events.

The large change in return period for seemingly small changes in rainfall amount mean that return level estimates appear less sensitive to the choice of distribution than return periods do. A spatial average rainfall of 48.09mm over 12-months, October-September, in the Northumbrian region is associated with a 250 year return period according to Pearson Type III, while for the GEV the 1 in 250 year event is only slightly larger, at 50.07mm. According to the Pearson Type III rainfall estimates of 45.83 and 44.56mm correspond to 1 in 500 and 750 year events, respectively. The corresponding GEV return level estimates are 48.36 and 47.42mm of rainfall.

However, it is important to note that the relatively high sensitivity of return period estimates is in part due to how it is calculated. Return periods are the reciprocal of the probability of exceedance. Therefore, the GEV probability of exceedances for Pearson Type III 500 and 750 year return levels (0.002 and 0.0013 probability of exceedance) are 0.00064 and 0.00034, respectively. The differences in probability of exceedance are 0.0014 and 0.001. The differences in return periods are 1068 and 2203 years, respectively. Between October 1958 and September 1959 an average of 45.6mm of rain fell in each month in the Northumbrian region. Figure ***14*** shows the difference in return period estimates for this event for the two distributions. According to Pearson Type III, this was a 1 in 695 year event, whereas according to the GEV it was a 1 in 1,755 year event. These are large differences which should be borne in mind for practical applications – although in practice, uncertainty will always be very high for such extreme events in the tails of the distribution. This is evidenced by the difference in plotting position of the two lowest magnitude rainfall events, and the associated distribution estimated return periods (Figure ***14***). This may be due to the uncertainty associated with plotting position estimates and/or distribution derived estimates. However, it is also possible these two points are the result of a separate process to that driving the distribution we see in the rest of the data. If determined to be the case, Peak-Over-Threshold (POT) may be more appropriate, as this would separate data below the threshold from the rest of the data. However, adopting a POT approach would limit the whole-rainfall regime focus and national applicability that is core to this approach.

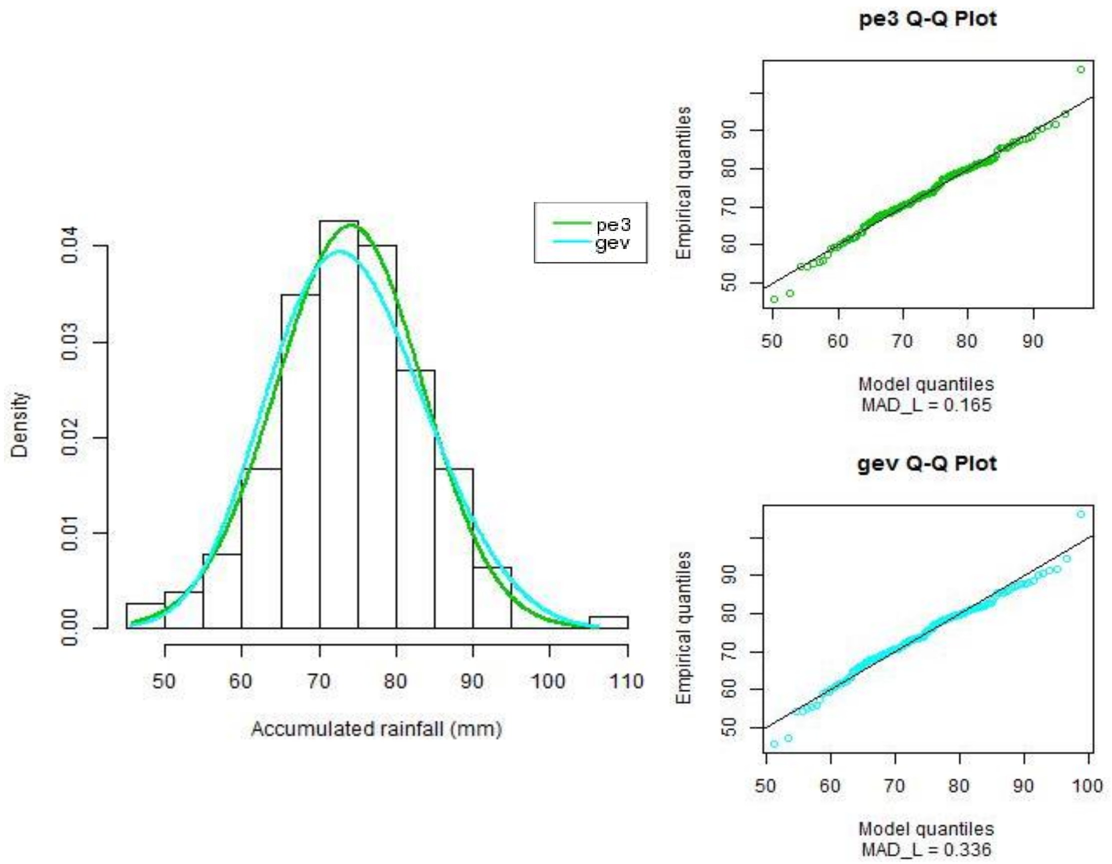


Figure 13 *Northumbrian Region 12-month accumulation period beginning in October: histogram with overlaid density functions of fitted Pearson Type III (pe3) and Generalised Extreme Value (GEV) distributions, and the corresponding QQ plots.*

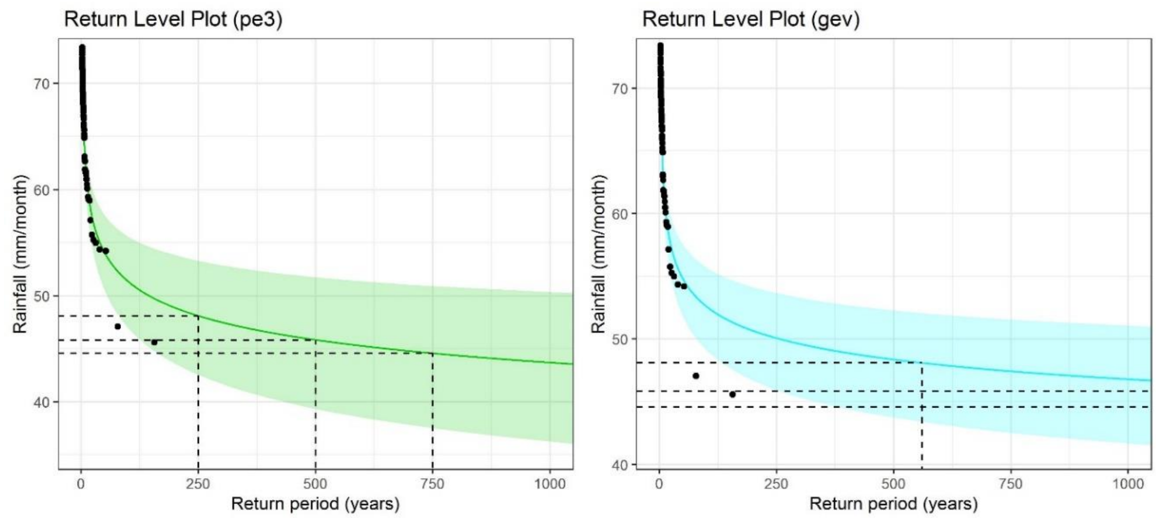


Figure 14 *Northumbrian 12-month accumulation period beginning in October: return period plots for The Pearson Type III (pe3) and Generalised Extreme Value (GEV) distributions, with associated 95% confidence interval bands, and highlighted return periods corresponding to 48.09, 45.83, and 44.56 mm of rainfall per month. Points correspond to the Weibull plotting positions of the observed rainfall accumulation data.*

5 Hydrological Summary Case Study

The UK Centre for Ecology & Hydrology (UKCEH) and the British Geological Survey (BGS) co-developed the National Hydrological Monitoring Programme in 1988 to monitor the hydrological conditions in the UK, and identify and interpret hydrological trends. This led to the development of the Hydrological Summary.

The Hydrological Summary, which is published monthly, documents the hydrological conditions in the UK, and places them in a historical context. This includes rainfall, river flow, groundwater, and reservoir data. Rainfall data are derived from 5km resolution gridded data, acquired from the Met Office National Climate Information Centre (NCIC). These data are grouped into large-scale hydrological regions, which for consistency through time are different from current measuring authority boundaries (in effect they are aligned to those used by the National Rivers Authority, pre-dating the Environment Agency and Natural Resources Wales, and earlier incarnations of Scottish Environment Protection Agency boundaries). Regional rainfall data are compared with 1981-2010 averages to enable the estimations of deviations from ‘normal’. Return periods are also estimated to quantify the rarity of four accumulation periods.

One of the aims of this project was to improve the current procedure employed to produce the hydrological summary return periods. Therefore, the Pearson Type III distribution was fitted to monthly rainfall data (averaged daily rainfall over the month). Due to the challenges with using a distribution without a lower bound at zero, for events that approach zero, the Gamma distribution is recommended for estimating return periods of short accumulation periods of 1-3 months. However, as the shortest accumulation period used for return period estimation in this particular case study was 6 months, the Pearson Type III distribution was used for all return period estimates.

We used the March 2012 Hydrological Summary for this case study. March 2012 was selected because there were above normal rainfall conditions in the North of the UK, while the South experienced below normal conditions. This allows the demonstration of the impact of the updated method on both the high and low tails of the distribution.

Return periods were estimated according to three methods, and are presented only in the form of intervals (Table 2), as this is how the return periods are currently presented in the Hydrological Summary. The Summary uses intervals, rather than the precise return period estimate, to acknowledge the uncertainty in the estimates.

Method 1: The current Hydrological Summary method

In Table 2, *HS 2012* refers to estimates using the current Hydrological Summary method. Rainfall data were accumulated for 6-, 13-, 18-, and 24-month periods, with respective start dates of October, March, October, and April. The Generalised Gaussian distribution (Griffin 2018) was fitted to the data, with parameters estimated using the method of moments. Return periods were then estimated according to the multiplicative inverse of the exceedance probability:

$$T = 1/p$$

where p is estimated using the Generalised Gaussian cumulative distribution function (Griffin 2018). Finally, uncertainty bands are assigned to each return period estimate, and presented in Table ***3***. Although the upper and lower limits of these bands vary in a consistent manner with the magnitude of the return period estimate, there is no clear rule for how they are derived, and in this sense the uncertainty bands can be considered arbitrary.

Method 2: The Hydrological Summary method, but with a different distribution

Return period bands were also estimated according to a slightly modified method, to determine the impact of distribution selection. Pearson Type III parameters were estimated according to the method of moments, and the Pearson Type III cumulative distribution function used to estimate the exceedance probability. All other methods were identical with the *HS 2012* method. These results are presented in Table ***3*** as *HS (pe3).*

Method 3: The proposed new method

Finally, to demonstrate the impact of switching to the approach outlined in this report, return period bands were estimated according to the methodology documented in Section 3. Non-overlapping time series were extracted from the Met Office NCIC data. Multiple time series were extracted for accumulation periods exceeding 12 months according to the process described in Section 2.1.1. Pearson Type III distributions were then fitted to each of the individual time series using maximum likelihood estimation. The bootstrapping process outlined in Section 3.2.1 was performed to extract the 50% confidence intervals (interquartile range), which was used in place of the arbitrary bands as an estimate of the uncertainty. These uncertainty bands are presented in Table 3 as *New.* Finally, the return periods and interquartile ranges were averaged over the constituent non-overlapping time series to obtain a single estimate for each accumulation period exceeding 12 months.

The estimation of uncertainty bands enabled improved characterisation of long return periods. In the current, and *HS (pe3)* methodologies, return period estimates exceeding 500 years are assigned the longest category, >>100. This is replaced in the *New* methodology, with a series of rules. The return period is estimated to be >>250 if the 25th percentile (Q25) of the return periods from the bootstrap is above 250 years. If the 75th percentile (Q75) is above 250, but Q25 is below 250, the return period is estimated to be >>Q25 (i.e. the Q25 estimated for that particular case).

Results

As shown in Section 4.1, more Pearson Type III null hypotheses were accepted by the goodness-of-fit tests than any other distribution investigated in this study. These results were consistent across accumulation periods, start months, and regions. Therefore, the goodness-of-fit of the distributions to the particular dataset used in the March 2012 Hydrological Summary is not evaluated further here. Instead, the difference in return period estimates is compared between the original method (fitting a Generalised Gaussian distribution using the method of moments), and two new methods (Pearson Type III – *HS (pe3)*, and Pearson Type III with updated processing and parameter estimation – *New*), to show the impact that an update of the Hydrological Summary methodology may have.

Merely replacing the Generalised Gaussian with the Pearson Type III distribution resulted in a different return period band estimation in 29 of 72 cases (Table 3). Of these, 17 estimates were higher than the Generalised Gaussian estimates, resulting in 12 cases where Pearson Type III produced lower return period band estimates than the Generalised Gaussian.

However, the difference in return period estimates between the Hydrological Summary (*HS 2012*) and Pearson Type III (*HS (pe3)*) methodology is relatively minor compared to differences between *HS 2012* and the updated methodology (*New*). The replacement of the Hydrological Summary methodology with the updated methodology had considerable effects on return period estimates. This is made evident by the lack of concurrence between the HS 2012 and HS (pe3) return period estimates presented in Table 3. The difference in estimates varies according to accumulation period and region (Figure 15). Whilst the differences in return period estimates typically increases with accumulation period, there is no clear spatial pattern, with considerable differences between return period estimates observed across the country. Comparison of the two heatmaps presented in Figure 15 demonstrates that larger differences also typically occur in more extreme events. Therefore, the apparent relationship between return period estimate difference and accumulation period may instead be a product of rarer events having larger differences.

Table 3. *Rainfall accumulations and return period estimates using the Hydrological Summary method and two updated methods.*

Area	**Rainfall**	**Mar 2012**	**Oct11-Mar12**				**Mar11-Mar12**				**Oct10-Mar12**				**Apr10-Mar12**			
				RP (years)				***RP (years)***				***RP (years)***				***RP (years)***		
				HS 2012	***HS (pe3)***	***New***		***HS 2012***	***HS (pe3)***	***New***		***HS 2012***	***HS (pe3)***	***New***		***HS 2012***	***HS (pe3)***	***New***
North West	mm	29	643	2-5	2-5	2-3	1293	2-5	2-5	4-5	1856	2-5	2-5	4-5	2352	2-5	2-5	5-6
	%	29	93				101				99				100			
Northumbria	mm	17	323	15-25	15-25	16-27	793	5-10	5-10	14-23	1254	2-5	2-5	7-11	1595	5-10	5-10	9-13
	%	25	71				88				98				96			
Midlands	mm	23	286	20-30	15-25	20-34	543	>100	>100	>>185	823	>100	>100	>>250	1140	80-120	>100	>>203
	%	39	70				66				70				75			
Yorkshire	mm	22	349	5-10	5-10	8-11	688	15-25	20-30	35-80	1078	10-15	10-20	22-41	1384	15-20	15-25	27-52
	%	33	77				78				85				85			
Anglian	mm	32	193	25-40	25-40	27-53	421	>100	>100	>>225	655	70-100	70-100	>>115	944	30-50	30-50	65-228
	%	70	63				65				72				78			
Thames	mm	25	235	20-35	25-40	23-42	512	40-60	50-80	>>74	794	50-70	60-90	>>97	1060	40-60	50-80	>>95
	%	46	62				68				74				76			
Southern	mm	29	283	20-35	20-35	21-38	571	80-120	80-120	>>232	959	25-40	20-35	46-145	1222	30-50	25-40	>>67
	%	49	62				68				78				78			
Wessex	mm	28	328	20-30	20-30	19-33	680	40-60	30-50	>>76	1029	70-100	80-120	>>122	1326	80-120	70-100	>>171
	%	41	65				73				75				77			
South West	mm	31	543	10-15	10-15	11-16	968	30-50	40-60	>>250	1484	50-80	70-100	>>250	1907	40-70	60-90	>>250
	%	32	72				74				76				79			
Welsh	mm	31	621	8-12	8-12	8-12	1169	10-20	10-20	38-109	1730	25-40	20-35	47-128	2271	15-25	15-25	35-83
	%	29	77				82				81				86			
Highland	mm	79	1309	10-20	10-20	13-20	2399	60-90	50-80	>>143	3115	10-15	10-15	19-32	3801	10-15	8-12	18-31
	%	49	120				128				111				111			
North East	mm	16	428	10-15	10-15	10-14	1096	2-5	2-5	4-5	1574	2-5	2-5	4-5	2102	2-5	2-5	5-7
	%	20	80				107				106				111			

Tay	mm %	34 29	726 92	2-5	2-5	2-3	1617 117	15-25	10-20	27-56	2295 112	10-15	8-12	16-26	2847 112	10-15	8-12	18-32
Forth	mm %	39 38	654 97	2-5	2-5	2	1430 116	10-15	10-20	25-48	2052 113	10-15	10-20	45-171	2539 112	8-12	8-12	20-36
Tweed	mm %	26 32	470 86	2-5	2-5	4-5	1138 110	2-5	2-5	6-8	1676 112	5-10	5-10	10-14	2067 108	2-5	2-5	5-6
Solway	mm %	52 43	960 112	5-10	5-10	7-10	1840 120	30-50	30-50	>>68	2591 114	20-30	15-25	>>250	3164 112	10-20	10-20	33-77
Clyde	mm %	74 46	1343 125	25-40	40-60	32-65	2440 129	>100	>100	>>250	3286 117	30-40	30-50	67-220	3960 114	15-25	20-35	>>63
Northern Ireland	mm %	22 23	694 108	5-10	5-10	6-8	1264 105	2-5	2-5	7-10	1749 100	2-5	2-5	4-5	2257 102	2-5	2-5	4-5

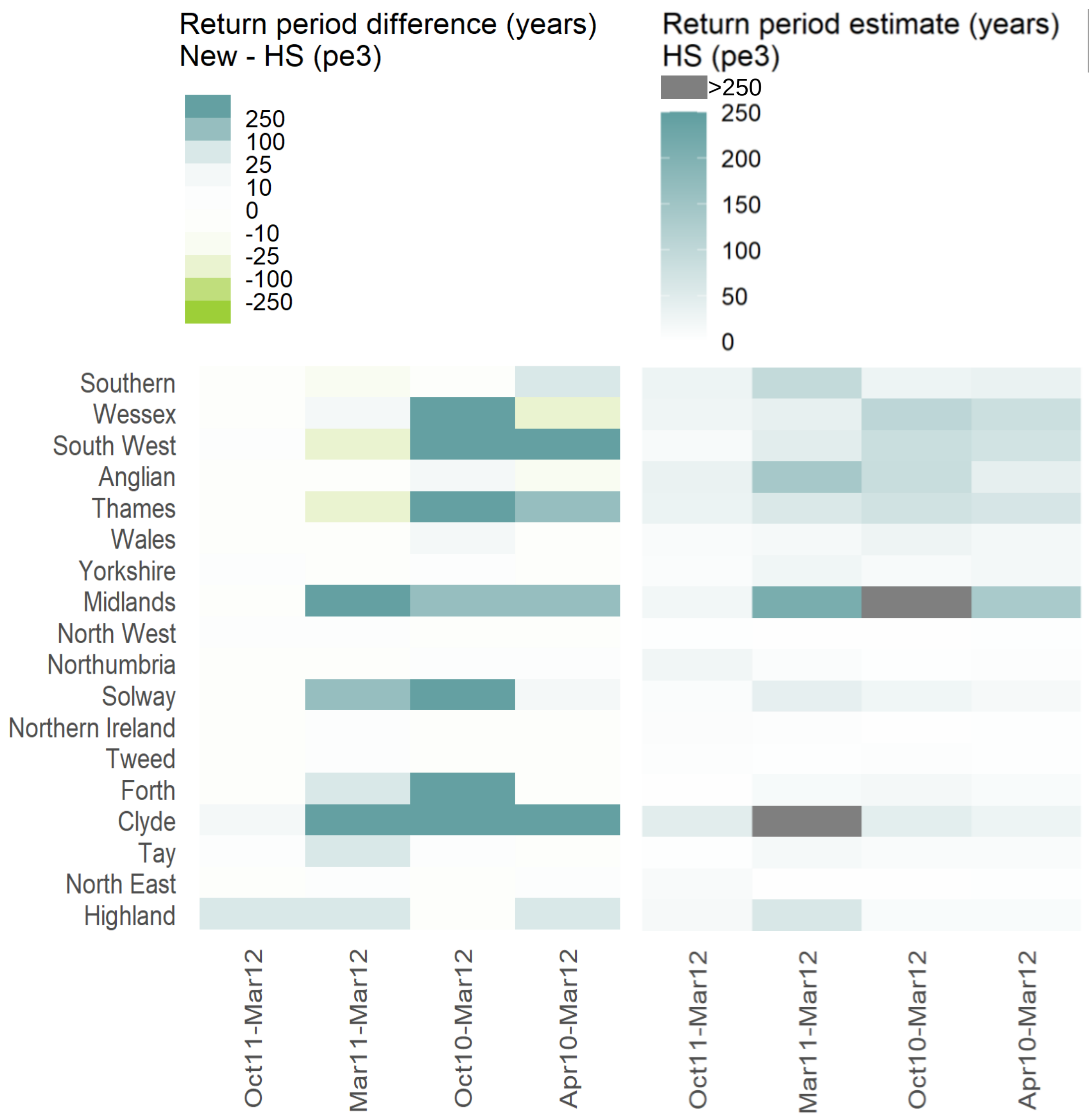


Figure 15 *Return period estimate from the HS (pe3) method, and difference in return period estimates between the New and HS (pe3) methods described above.*

The replacement of the arbitrary return period bands with the new boostrapped interquartile range resulted in larger uncertainty ranges for long return periods. This demonstrates the difficulty in accurately estimating the return period of rare extreme events. In contrast, short return period estimates exhibited relatively narrow uncertainty bands compared with the original method. This provides evidence-based support for confidence in precise short return period estimates, while estimates of long return periods are more uncertain.

Figure ***16*** visualises an example of the differences exhibited in Table 3. A visual inspection of Figure ***16*** suggests the differences in the median return period estimates derived from the “New” method, and the return period estimates derived from the original “HS 2012” method are minimal. However, when investigating actual return level magnitudes, the original methodology underestimates the 250, 500 and 750 year return levels compared with the updated methodology. Furthermore, the stepped change in uncertainty band estimates is removed in the updated methodology, producing a more realistic, evidence-based estimate of uncertainty.

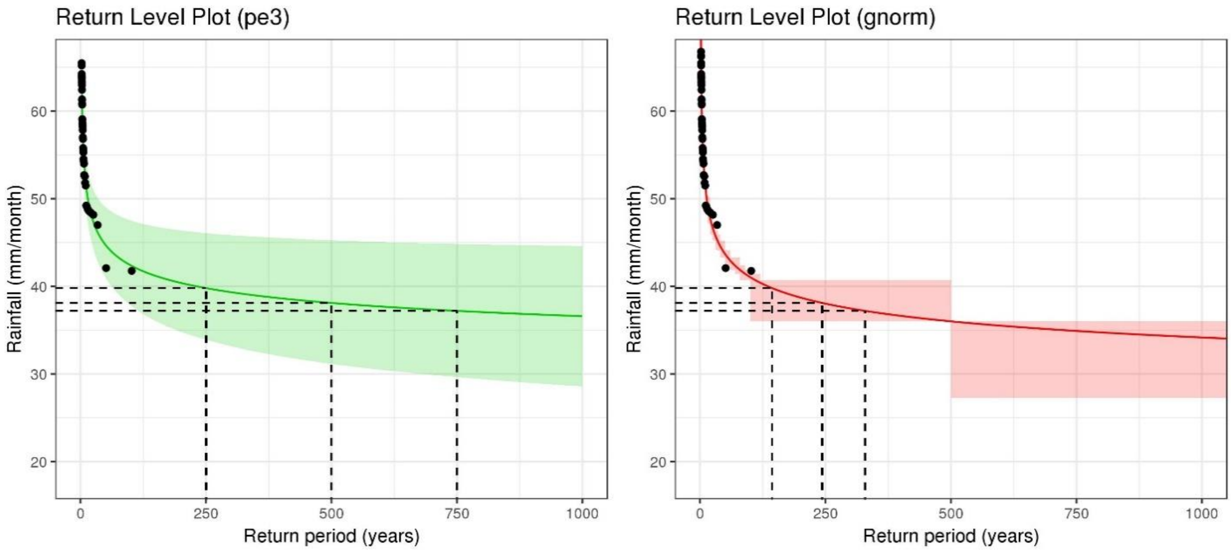


Figure 16 *Return periods and corresponding return levels for March-October (6-months) rainfall accumulations in the Midlands. Plots vary in their estimation methodology (left – updated ("New"); right – original ("HS 2012")). Dashed lines correspond to the 250, 500 and 750 year return levels estimated using the updated methodology.*

6 Summary

The results provided in this report provide a foundation for improved return period estimation for the Hydrological Summary and the Drought Libraries. They could provide useful guidance more generally for appraising the rarity of n-month rainfall accumulations for the assessment of drought severity. However, it is important to note that this is only one approach to drought severity assessment, and other methods could yield different results for the same drought event. In particular, our approach focuses on n-month accumulations with a defined start or end month – return period assessments for n-month accumulations starting in any month would inevitably be lower (Reed 2011). Moreover, there are many definitions of drought and correspondingly different indicators used for drought characterisation (e.g. Bachmair *et al.* 2016). Use of other indicators (such as the Standardised Precipitation Index, SPI, e.g. Svensson *et al*. (2016) for UK applications) to quantify meteorological drought severity could also lead to different outcomes. Similarly, here our focus is exclusively on meteorological drought and, more narrowly, precipitation, rather than evaporative demand, as is included in some indicators such as the Standardised Precipitation and Evapotranspiration Index (SPEI, Vicente-Serrano *et al.* 2010) or other Aridity Indices (Marsh *et al.* 2007). Return Period assessments incorporating evapotranspiration, or based on hydrological indicators (river flow, groundwater) or metrics of water supply system performance, could lead to different results. Finally, and in keeping with the need for the Hydrological Summary to characterise the whole range of variability, we note that here we have focused on applying probability distributions to whole time series of n-month rainfall, incorporating wet as well as dry extremes. Other approaches to return period assessment take a different approach and apply a peak-over-threshold model to focus on the dry extremes (e.g. Burke *et al.* 2010).

However, the results presented here demonstrate potential to inform best practice when estimating drought return periods for durations up to 5 years in the UK. The Pearson Type III distribution was accepted (using the Modified-Anderson-Darling test) more often than any of the other probability distributions when grouping all rainfall data together. This supports the use of the Pearson Type III distribution for drought return period estimation in the UK. Whilst the null acceptance rate of the Pearson Type III distribution varied across the country and over the different accumulation periods, it remained the most appropriate distribution for estimating the rarity of long-duration rainfall in the UK.

The Gamma distribution, the zero-bounded version of the Pearson Type III distribution, was deemed more appropriate for short accumulation periods due to its ability to fit the observed data and prevent the production of negative return levels. Overall, the Gamma, Generalised Logistic and Generalised Extreme Value distributions also exhibited relatively high acceptance rates. This was paired with certain distributions outperforming the Pearson Type III in specific cases. However, there is a benefit to adopting a nationally consistent approach, particularly in bringing simplicity and facilitating inter-comparisons between regions for the purpose of improved national planning. Therefore, the consistency of performance of the Pearson Type III distribution across all accumulation periods, start months and regions, and the Gamma distribution for 1 to 3-month accumulation periods, supports the adoption of Pearson Type III for estimating the frequency of low-magnitude rainfall events exceeding 3-month accumulation periods, and the Gamma distribution for the frequency estimation of 1 to 3 month low-magnitude accumulated rainfall events.

7 Acknowledgements

The authors would like to thank Richard Davis and Ralph Ledbetter of the Environment Agency for their support developing and reviewing this research and subsequent report, as well as Sandie Clemas (UKCEH) for her helpful comments on a draft. This work was supported by the Natural Environment Research Council (NERC) via two awards: (1) award number NE/R016429/1 (funding for the UK National Hydrological Monitoring Programme as part of the UK-SCAPE programme delivering National Capability); and (2) award number NE/L01061X/1, ‘Engaging diverse stakeholders and publics with outputs from the UK Drought and Water Scarcity Programme’ (ENDOWS).

8 Reference List

Anderson, T. W.; Darling, D. A. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. (1952). *Annals of Mathematical Statistics, 23*(2), 193-212. https://doi.org/10.2307%2F2286009

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., Knutson, C., Smith, K.H., Wall, N., Fuchs, B., Crossman, N.D. and Overton, I.C. (2016). Drought indicators revisited: the need for a wider consideration of environment and society. *WIREs Water*, *3*(4), 516-536. https://doi.org/10.1002/wat2.1154

Burke, E. J., Perry, R. H. J., Brown, S. J. (2010). An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology, 338*(1-2), 131-143. https://doi.org/10.1016/j.jhydrol.2010.04.035

Counsell, C., Hunt, D. and Ledbetter, R. (2017) Drought Vulnerability Framework. UKWIR report Ref: 17/WR/02/12. UKWIR, London, 125pp. https://ukwir.org/drought-vulnerability-framework-0.

Environment Agency. (2020) Meeting our Future Water Needs: A National Framework for Water Resources. Version 1. Bristol: Environment Agency. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872759/National_Framework_for_water_resources_main_report.pdf

Griffin, M. (2018) gnorm: Generalized Normal/Exponential Power Distribution. R package version 1.0.0. https://CRAN.R-project.org/package=gnorm

Gudmundsson, L. and Stagge, J.H. (2016) SCI: Standardized Climate Indices such as SPI, SRI or SPEI. R package version 1.0-2. https://CRAN.R-project.org/package=SCI

Guillod, B. P., Jones, R. G., Bowery, A., Haustein, K., Massey, N. R., Mitchell, D. M., Otto, F. E. L., Sparrow, S. N., Uhe, P., Wallom, D. C. H., Wilson, S., and Allen, M. R. (2017). weather@home 2: validation of an improved global–regional climate modelling system, Geosci. Model Dev., 10, 1849–1872, https://doi.org/10.5194/gmd-10-1849-2017.

Hollis, D., McCarthy, M., Kendon, M., Legg, T. and Simpson, I. (2019) HadUK-Grid – A new UK dataset of gridded climate observations. *Geoscience Data Journal.* https://doi.org/10.1002/gdj3.78.

Kharin, V.V. and Zwiers, F.W. (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *Journal of Climate, 13*, 3760–3788. https://doi.org/10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2.

Kyselý, J. (2007) A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *Journal of Applied Meteorology and Climatology, 47*, 3236-3251. https://doi.org/10.1175/2008JAMC1763.1.

Mann, H.B. (1945) Nonparametric tests against trend. *Econometrica*, 13, 245-259.

Marsh, T., Cole, G. and Wilby, R. (2007) Major droughts in England and Wales, 1800-2006. *Weather*, *62*, 87-93. https://doi.org/10.1002/wea.67.

McLeod, A.I. (2011) Kendall. R package version 2.2. https://cran.r-project.org/web/packages/Kendall

Paeth, H. and Hense, A. (2005) Mean versus extreme climate in the Mediterranean region and its sensitivity to future global warming conditions. *Meteorologische Zeitschrift, 14*, 329-347. https://doi.org/10.1127/0941-2948/2005/0036.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reed, D.W. (2011) 'Tabony tables' in *Letters in applied hydrology.* Wallingford, UK: DWRconsult.

Sinclair, C.D., Spurr, B.D. and Ahmad, M. I. (1987) Modified Anderson Darling test, technical report, Dep. of Math. Sci., St. Andrews Univ., St. Andrews, Fife, Scotland.

Sinclair, C.D., Spurr, B.D. and Ahmad, M.I. (1990) Modified Anderson Darling test. *Communications in Statistics – Theory and Methods, 19*(10), 3677-3686, https://doi.org/10.1080/03610929008830405.

Stephenson, A.G. (2002) evd: Functions for Extreme Value Distributions. R package version 2.3-3. https://CRAN.R-project.org/package=evd

Svennson, C., Brookshaw, A., Scaife, A.A., Bell, V.A., Mackay, J.D., Jackson, C.R., Davies, H.N., Arribas, A. and Williams, A. (2015) Long range forecasts of UK winter river flow. *Environmental Research Letters,* 10 (6), 064006 (8 pp.). https://iopscience.iop.org/article/10.1088/1748-9326/10/6/064006.

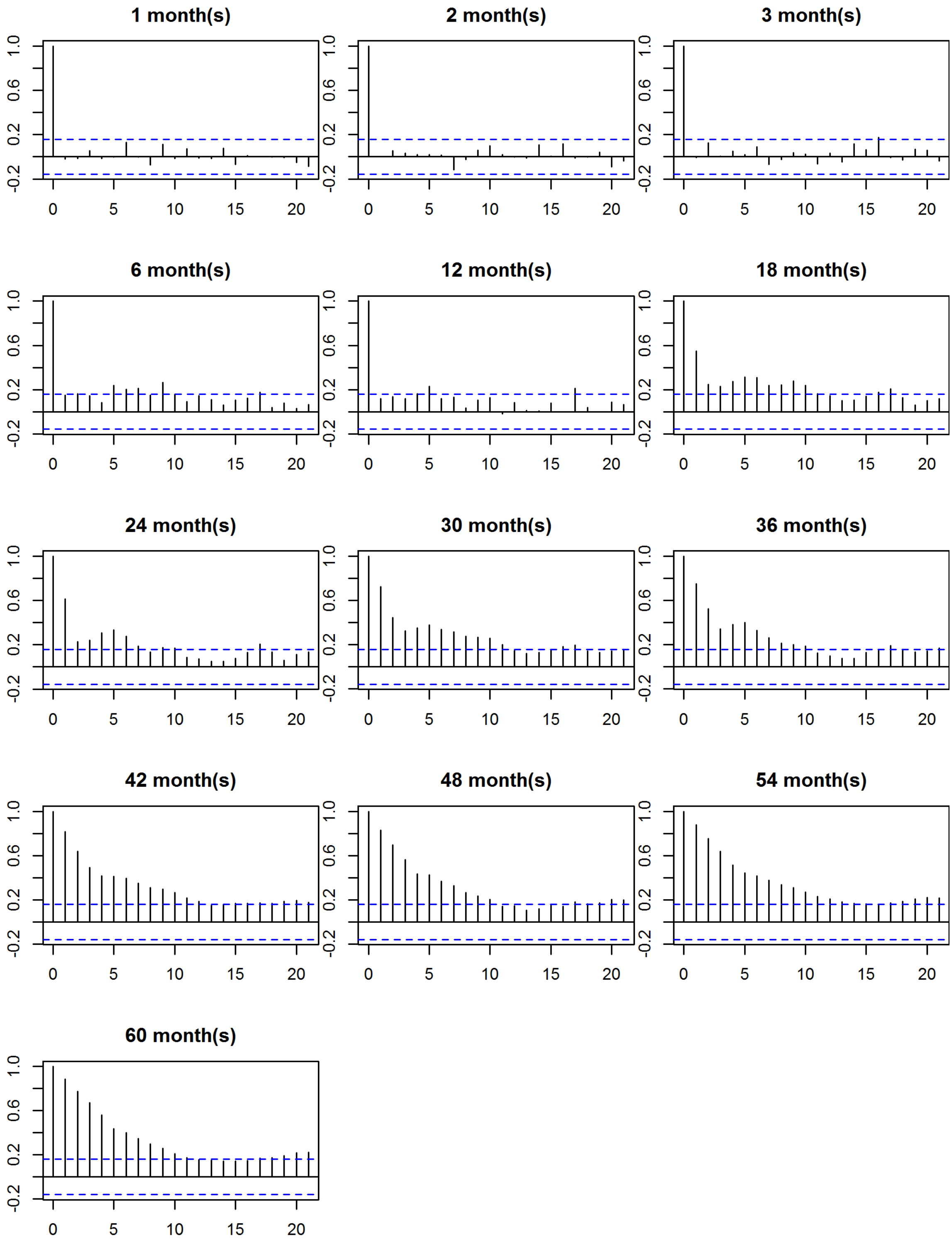
Svensson, C., Hannaford, J. and Prosdocimi, I. (2016) Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications. *Water Resources Research, 53*, 999-1018. https://doi.org/10.1002/2016WR019276.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer-Verlag.

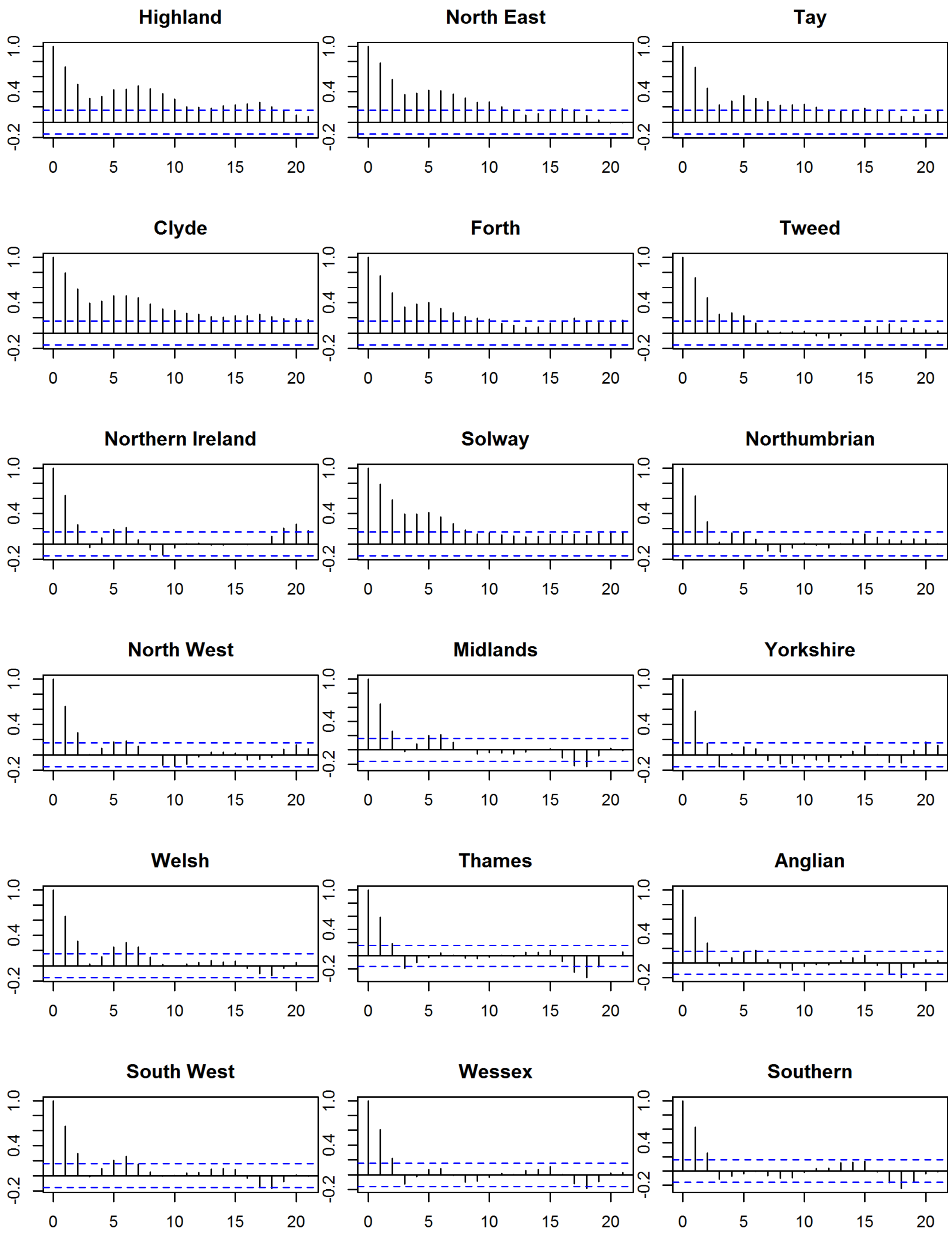
Vicente-Serrano, S.M. (2010) A multiscalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23,* 1696-1718. https://doi.org/10.1175/2009JCLI2909.1.

9 Appendices

9.1 Appendix 1 – Autocorrelation plots of Forth accumulated rainfall events beginning in October, varying by accumulation length. The time series consist of overlapping events for durations of 12 months or more. The x-axis shows units of events.



9.2 Appendix 2 – Autocorrelation plots of 36-month (overlapping) accumulated rainfall events beginning in October, in each of the 18 regions. The x-axis shows units of events, that is, in this case a 3-year period.



INVESTORS IN PEOPLE

Athena SWAN Bronze Award

PEER

BANGOR
UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW
United Kingdom
T: +44 (0)1248 374500
F: +44 (0)1248 362133

EDINBURGH
UK Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB
United Kingdom
T: +44 (0)131 4454343
F: +44 (0)131 4453943

LANCASTER
UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP
United Kingdom
T: +44 (0)1524 595800
F: +44 (0)1524 61536

WALLINGFORD (Headquarters)
UK Centre for Ecology & Hydrology
Maclean Building
Benson Lane
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB
United Kingdom
T: +44 (0)1491 838800
F: +44 (0)1491 692424

enquiries@ceh.ac.uk

www.ceh.ac.uk